

Comments of the Internet Archive on the U.S. Copyright Office Notice of Inquiry on the Digital Millennium Copyright Act Section 512 Safe Harbors

March 22, 2016

Introduction

The Internet Archive thanks you for this opportunity to comment on the DMCA safe harbors.

The Internet Archive is a 501(c)(3) non-profit organization based in San Francisco, California. Our mission is to provide universal access to all human knowledge. As part of that mission, we collect, archive, and provide public access to many different types of material digitally, including websites, music, software, images, books, educational materials, video games, films, ephemera, and more. Some of these materials we collect ourselves. However, many of the materials in our collections were uploaded by third-party users of the Internet Archive—librarians, archivists, enthusiasts, collectors, and other members of the public. The DMCA safe harbors help us do this. Given the high statutory damages permitted by the Copyright Act and high court costs, a copyright infringement lawsuit could be enough to cripple our small nonprofit organization.

We provide comments here in our capacity as both as an online service provider that hosts so-called “user generated content” and as a library with a mission to preserve and provide public access to cultural materials. As we move increasingly towards a world where human knowledge is stored digitally, we are likely to see more libraries playing the role of host and curator of content posted by users. As such, it is important to understand how library interests intersect with the DMCA safe harbors and to ensure that libraries continue to enjoy the protection of these safe harbors in the future.

Relevant Questions from the Notice of Inquiry

We respond below to the questions where our unique position at the intersection of Internet and the library world may be able to shed light for the Copyright Office into the ways in which the DMCA safe harbors are working well, and allow us to address areas for potential improvement.

Question 5. Do the section 512 safe harbors strike the correct balance between copyright owners and online service providers?

In crafting the DMCA, Congress created a system of shared responsibility for managing potential copyright infringement online. On the whole, we believe that this system is working well, and should not be significantly overhauled. The DMCA safe harbors provide important certainty, allowing us to collaboratively build our collections with our community. The DMCA safe harbor has allowed many online communities, such as the Internet Archive’s, to grow and thrive. Our community regularly contributes older, at-risk materials for preservation and public access. For example, our some of our community collections

include feature length films¹, short films², old radio programs³, early 20th Century 78rpm records and cylinder recordings⁴, and pre-1964 architectural trade catalogs, house plan books, and technical building guides that document past design and construction practices⁵. Without the protection of the DMCA safe harbors, we might not be able to host collections like these—despite the fact that no one has complained about the vast majority of the materials.

There are significant burdens on both sides of the DMCA notice and takedown process. The DMCA places the burden of identifying and notifying service providers of claimed infringement squarely on the owners of the copyrighted material. Upon notice, the responsibility for removing or disabling access to that material shifts to the service provider. This balance makes sense, as it places the burden on the party with the relevant knowledge and ability to act. For example, only copyright holders know what they own, what materials they have licensed, and where their materials are allowed to appear. Given that many of the works in our collections are older and have no current commercial life, it can be very difficult to determine who the owner may be, or whether they are still protected by copyright at all. Many users place works in our collections (as opposed to placing them on a commercial platform) in order to ensure that they are preserved for future generations—with no intent to violate the law.

The DMCA’s express provision that service providers have no affirmative duty to monitor for infringing activity remains an extremely important safeguard both for free speech and for the continuation of traditional library activities in the digital age. There is a distinction between commercial piracy and noncommercial preservation and sharing of our cultural heritage. The context in which a user posts material that is owned by another person or entity must be evaluated before determining whether such posting is infringement or fair use. This is why proposals for “notice and staydown,” which would appear to require platforms to use automated processes to make sure certain materials are never again able to be posted to the internet—regardless of context—threaten to chill legitimate speech and fair uses of materials. Libraries have long been champions of intellectual freedom, and are encouraged by the Library Bill of Rights to “challenge censorship in the fulfillment of their responsibility to provide information and enlightenment.”⁶ A “notice and staydown” regime would violate these fundamental principles.

The takedown process may seem trivial in the era of YouTube’s automated Content ID system. But the vast majority of service providers do not have the resources to develop such technology, and instead rely on human review and responses to notices of claimed infringement. We, for example, have no in-house legal staff and our overall staffing is very small in relation to the amount of data we are able to host. Considering that we have over 26 petabytes of data in our collections, and about 100,000 items per month are uploaded by users other than Internet Archive staff, we receive a relatively small number of DMCA

¹ See https://archive.org/details/feature_films

² See https://archive.org/details/short_films

³ See <https://archive.org/details/oldtimeradio>

⁴ See <https://archive.org/details/78rpm>

⁵ See <https://archive.org/details/buildingtechnologyheritagelibrary>

⁶ The Library Bill of Rights, available at:

<http://www.ala.org/advocacy/sites/ala.org.advocacy/files/content/intfreedom/librarybill/lbor.pdf>.

notices. Nevertheless, we take our role seriously, and we devote a significant amount of time and resources to dealing with the notices we do get. As described further below, some notices take much longer than others to process.

The current system already imposes significant burdens on service providers, and we believe that altering these burdens to shift more responsibility to affirmatively monitor for infringement would force small nonprofits and libraries without huge budgets for legal staff to divert even more of their already strained resources into managing copyright claims, or else carry huge risk of liability.

Question 8. In what ways does the process work differently for individuals, small scale entities and/or large scale entities that are sending and/or receiving takedown notices?

As the Copyright Office has recognized, many different types of entities currently enjoy safe harbor protection, including nonprofit libraries. Recent advances in technology have allowed libraries to house increasingly massive collections of electronic data, both through their own digitization efforts, and through community collection building that relies on materials posted by other parties. The DMCA offers much needed legal protection and certainty for these sorts of activities. Libraries and other non-profit organizations are unlikely to be able to bring to bear the sorts of resources that larger commercial entities may have access to in terms of staff and technological tools for automated or more efficient processing of copyright claims. Having fewer automated tools and a smaller workforce means we may not always have ability to:

- take down offending content with precision;
- process as rapidly as bigger companies; or
- closely review claims with any degree of complexity.

Any proposed changes to the burdens associated with reviewing and removing content under the DMCA should take these issues into consideration.

Question 9. Please address the role of both “human” and automated notice and takedown processes under section 512, including their respective feasibility, benefits, and limitations.

We receive DMCA notices with some issue that requires clarification at least every week. Processing such a request might require reviewing each individual item identified, composing responses requesting clarification and/or clarifying when use is limited (e.g., lending/print-disabled access), requesting information for incomplete notices, processing and confirming proper take down of the item, and sending notice to the uploader and claimant after the item has been taken down. This clarification process takes thoughtful human review, and is not something that could easily be automated without the risk of many works being improperly removed from our collections.

From what we are able to tell, a large proportion of the improper or incomplete notices we receive appear to come from third-party companies on behalf of major studios or

publishers. A number of these third-party services routinely send improper notices. Some examples of the types of improper notices we have received include:

- Notices that mistakenly identify works that are in the public domain. For example, we have received notices that mistake volunteer audio recordings of classic works such as *Jane Eyre*, *Sense and Sensibility*, Bram Stoker's *Dracula*, *Moby Dick*, and *Little Women* for commercial audiobook editions.
- Notices that use loose keyword matching that overclaims works that are clearly not owned by the major content holders they represent. For example, we received a takedown notice regarding an old Salem cigarette commercial based on the term "Salem" which is also the title of a major television series. Similar keyword misidentifications frequently show up as "matches" for music, concerts, home movies, and public domain books.
- Notices sent regarding reviews or lesson plans about a given work, rather than for the work itself. For example, we received a takedown notice regarding a lesson plan from the Department of Education about "To Kill a Mockingbird." Similarly, Warner Brothers has sent takedown notices for reviews of films and television programs mistaken for the works themselves.
- Notices containing malformed URLs that do not point to any existing materials on our system.
- Repeated notices for materials weeks or months after they have already been removed and notice of such takedown had been sent to the claimant.

These are just a few examples of the types of notices we receive that may require a fair amount of time to deal with properly.

We also routinely receive notices that are difficult to process because they do not specifically identify any works or creators, or they only identify creators without identifying any specific works. Other notices include additional and vague threatening language regarding rights other than copyright. For example, Web Sheriff's notices often include a clause that says: "Infringed Rights: COPYRIGHT / PERFORMERS' RIGHTS / MORAL RIGHTS / RIGHT-OF-PUBLICITY / PERSONAL GOODWILL & REPUTATION / BUSINESS GOODWILL & REPUTATION / CONSUMER PROTECTION RIGHTS as applicable"). We often get notices seeking to use the DMCA process to address trademark, privacy, and defamation, among other non-copyright issues.

As long as these inaccurate, improper and/or incomplete notices are sent to us, human review on our end of the process is required so as not to overly censor legitimate speech and online activity. Higher standards for automated takedown notices could reduce the burden on our small staff in having to clarify or otherwise process these notices.

Question 11. Are there technologies or processes that would improve the efficiency or effectiveness of the notice and takedown process?

While technology and automated processes can help the notice and takedown process to scale in certain ways, human review must remain a crucial part of the process on both

sides. Otherwise, mistakes can lead to censorship and chilled speech. In addition to the examples above, there are other situations where it is not clear that the use of the material claimed in a DMCA notice is infringing. For example, we have received notices for material for which there is clearly no commercial interest, instead, the claim appears to be directed at preventing embarrassment or silencing criticism (e.g., a claim on a picture used in a critical video). We have also received DMCA notices that apply to a very small portion of a larger work, for example, comments on an archived message board or website guestbook, a poem on the homepage of an archived literary journal website, a quote in a yearbook, or cd cover art that is mistakenly identified as the full album. We also receive frequent notices for materials that we have permission to archive, such as live concert recordings hosted at archive.org/details/etree. Processing such claims generally requires human review, and some amount of back and forth explanation and discussion with the copyright holder.

We are deeply concerned that automated filtering could lead to taking down many materials that are being used in reasonable, legitimate and legally protected ways—especially when the underlying purpose of the complaint is not copyright related but rather an attempt to silence critical speech.

Question 12. Does the notice and takedown process sufficiently protect against fraudulent, abusive, or unfounded notices? If not, what should be done to address this concern?

Given the high number of inaccurate, improper and/or incomplete notices we receive on a regular basis, it seems reasonable to conclude that this sort of behavior is not being properly disincentivized. There are currently no real penalties for sending overly broad and inaccurate notices, and no incentives for sending accurate, well-formatted notices. Under the current system, copyright holders have a unilateral weapon that allows them to send inaccurate, improper and/or incomplete notices in bulk, without repercussion. Many copyright owners act in good faith, but for those who do not, there should be real penalties to deter bad behavior.

Further, the law is structured to incentivize taking materials down, rather than leaving them up when the situation is unclear. Any time we receive a DMCA takedown notice and we decide not to take the material down for any reason, we risk our safe harbor protection. Users are not always in a position to be able to file a counter-notice since most would not have the resources to fight a legal battle in court should the copyright holder decide to file suit, even in cases where they would ultimately prevail. Some commercial platforms such as Google and Automattic have been able to stand up for their users in court, and we applaud them for doing so. However, many platforms—especially non-profit libraries—will never be in a position to be able to take bad copyright actors to court.

It might make sense to create a provision in the law that would grant the service provider the ability to refuse to take material down when they have a reasonable, good faith belief that the material identified in a DMCA notice is non-infringing. For example, if a work appears to be in the public domain, or if the use of the material appears to be a fair use, then the service provider could refuse to take the material down without risking the imposition of

statutory damages. In combination with a simple, inexpensive dispute resolution process in cases where the copyright holder disagrees with the service provider's decision, this could lead to far fewer bad notices being sent and fewer takedowns of legitimate materials.

Question 23. Is there sufficient clarity in the law as to what constitutes a repeat infringer policy for the purposes of section 512 safe harbors? If not, what should be done to address this concern?

The DMCA statute does not define what a "repeat infringer" is, but it conditions safe harbor protection on the service provider's ability to reasonably implement a policy of terminating the accounts of such infringers in appropriate circumstances. Congress left open not only the question of what a "repeat infringer" is, but also what "appropriate circumstances" are. As such, service providers must come up with their own definitions of these vague terms. This creates a fair amount of uncertainty, but also permits some flexibility which can have advantages.

Our community includes hundreds of volunteer archivists who actively seek to preserve at-risk websites, software, old audio recordings, home videos, and other older materials whose commercial life (if they ever had one) is long past. We have no problem whatsoever with taking down current commercially viable materials, and terminating the accounts of users who repeatedly or flagrantly post such materials. But we have no desire to unfairly punish users who either made a genuine mistake in circumstances where rights issues were unclear (e.g., we received notices after the *Golan v. Holder* decision removed certain materials from the public domain retroactively) or users who post materials in a manner that may be a fair use but choose for whatever reason not to contest a takedown notice.

In our community, we need to distinguish between those users who repeatedly upload current, commercially viable material from those who operate in good faith by preserving ephemera and other older materials for posterity. There is a difference between intentional commercial infringement and innocent noncommercial infringement. The copyright damages system takes this difference into account, ranging from "innocent" to "willful." We believe that a properly implemented repeat infringer policy must take these considerations into account as well, under the auspices of "appropriate circumstances."

Respectfully submitted,

The Internet Archive